



## Busca en Internet información sobre estas cuestiones:

- Busca éxitos o logros en el entorno del Procesamiento del Lenguaje Natural, fundamentalmente en el siglo pasado.

### **Turing Test (1950)**

- Descripción: Propuesto por Alan Turing en su artículo "Computing Machinery and Intelligence", el Turing Test se convirtió en un estándar para evaluar la inteligencia de una máquina basada en su capacidad para generar respuestas indistinguibles de las de un humano en una conversación.
- Importancia: Marcó el inicio del debate sobre la inteligencia artificial y el PLN, inspirando investigaciones futuras.

### **Machine Translation - IBM y Georgetown (1954)**

- Descripción: IBM y la Universidad de Georgetown colaboraron en el primer experimento significativo de traducción automática, traduciendo más de 60 oraciones del ruso al inglés.
- Importancia: Demostró el potencial de las computadoras para realizar tareas de lenguaje humano, estableciendo una base para la investigación en traducción automática.

### **Eliza (1966)**

- Descripción: Creado por Joseph Weizenbaum, Eliza fue uno de los primeros programas de procesamiento de lenguaje natural que podía mantener una conversación con un humano simulando un psicoterapeuta Rogeriano.
- Importancia: Aunque limitado, Eliza demostró la capacidad de los programas de computadora para interactuar con humanos en un lenguaje natural, popularizando el interés en la IA y el PLN.

### **WordNet (1985)**

- Descripción: Desarrollado por George Miller y su equipo en la Universidad de Princeton, WordNet es una base de datos léxica del inglés que agrupa palabras en conjuntos de sinónimos (synsets) y proporciona relaciones semánticas entre ellos.
- Importancia: Se convirtió en una herramienta crucial para diversas aplicaciones de PLN, incluyendo la desambiguación de palabras y la comprensión semántica.

### **Statistical Machine Translation (1990s)**

- Descripción: La investigación en traducción automática estadística, impulsada por investigadores como Peter Brown y su equipo en IBM, desarrolló métodos que utilizaban modelos probabilísticos para traducir texto.
- Importancia: Revolucionó el campo de la traducción automática, mejorando la calidad de las traducciones y sentando las bases para los sistemas modernos de traducción.



## Part-of-Speech Tagging (1990s)

- **Descripción:** Se desarrollaron métodos eficientes para el etiquetado de partes del discurso (POS tagging) utilizando técnicas como HMM y algoritmos de máxima entropía.
- **Importancia:** Estas técnicas mejoraron la precisión del análisis gramatical de texto, facilitando el desarrollo de aplicaciones avanzadas de PLN.

- Busca limitaciones e inconvenientes en el Procesamiento del Lenguaje Natural.

El PLN (procesamiento del lenguaje natural) tiene que enfrentarse a día de hoy a varias limitaciones e inconvenientes. Podemos destacar:

- **Ambigüedad del Lenguaje.**
  - **Polisemia y homonimia:** Las palabras pueden tener múltiples significados (polisemia) o pueden ser fonéticamente idénticas pero tener significados diferentes (homonimia).
  - **Ambigüedad Sintáctica:** Una oración puede tener múltiples estructuras gramaticales y necesitar una comprensión contextual profunda.
- **Contexto y Comprensión.**
  - **Comprensión del contexto:** Para entender el significado completo de un texto requiere un conocimiento contextual del mundo real, algo que a los modelos PLN les resulta difícil.
  - **Conocimiento del mundo:** Muchas tareas del PLN requieren conocimiento general del mundo, que no tiene porque estar presente en los datos de entrenamiento del modelo.
- **Datos y Sesgos**
  - **Dependencia de grandes cantidades de datos:** Se necesitan ingentes cantidades de datos etiquetados para entrenar los modelos PLN.
  - **Sesgos en los datos:** Los modelos pueden heredar y amplificar sesgos presentes en los datos de entrenamiento.
- **Interacción Humano-Maquina**
  - **Interacción Natural:** Resulta complejo lograr una interacción fluida y natural entre humanos y máquinas.
  - **Interpretabilidad y Transparencia:** Algunos modelos de PLN, especialmente los basados en aprendizaje profundo, son cajas negras, es decir que sus decisiones no son fácilmente interpretables.

- Busca características que debe cumplir un PLN moderno.

Un PLN moderno debe cumplir varias características:

- Debe ser preciso y robusto, es decir el modelo debe ser capaz de realizar las tareas específicas con precisión, tareas como la traducción automática o el reconocimiento de las entidades nombradas, etc... También debe manejar de manera robusta los errores tipográficos, las variaciones lingüísticas, las distintas jergas, etc...
- Debe ser escalable y eficiente.
- Debe ser capaz de dar soporte a múltiples lenguajes.



- Debe ser capaz de comprender el contexto y ser capaz de incorporar conocimiento del mundo real.
  - Debe ser capaz de aprender y adaptarse a múltiples situaciones.
  - Debe facilitar una interacción natural y fluida entre el modelo y los humanos.
  - Debe minimizar los sesgos para asegurar resultados justos y equitativos.
  - Debe asegurar la privacidad y la seguridad de los datos personales utilizados y generados por el modelo PLN.
  - Debe ser versátil y poder aplicarse a diversos dominios como puede ser la medicina, la educación, etc...
- Mantén una breve conversación con un chatbot (Eliza o Alizia), ¿cuál es tu impresión? ¿cómo crees que se podría mejorar?

Mi impresión con estos dos chatbots es que son un primer acercamiento en cuanto a modelos PLN, ya que me da la impresión de que no consiguen entender correctamente las diferentes conversaciones que hemos tenido, ya que a veces respondía como de forma genérica sin haber tenido en cuenta el contexto anterior de la conversación. Se podrían mejorar logrando por un lado que comprendan mejor el contexto de una conversación y dotándolas de mejoras en cuanto a la forma de interactuar entre humano y máquina.

- Nombra un par de aplicaciones PLN de clasificación de documentos y de análisis de sentimiento y opinión. Descríbelas brevemente.

#### Clasificación de Documentos.

- **Google News:** Puede analizar automáticamente los artículos de diferentes fuentes y clasificarlos en categorías pertinentes. Esto permite a los usuarios encontrar rápidamente noticias sobre temas de interés.
- **Evernote:** Utiliza tecnologías de PLN para ayudar a los usuarios a organizar sus notas y documentos.

#### Análisis de Sentimiento.

- **Amazon Comprehend:** Amazon Comprehend es un servicio de PLN de Amazon Web Services (AWS) que permite analizar el sentimiento de textos.
- **IBM Watson Natural Language Understanding:** Ofrece herramientas para analizar el sentimiento y la opinión de textos en redes sociales, reseñas u otros medios.